14th World Congress on Computational Mechanics (WCCM)
ECCOMAS Congress 2020
19–24 July 2020, Paris, France

NUMA-AWARE STRATEGIES FOR THE HETEROGENEOUS EXECUTION OF SPMV ON MODERN SUPERCOMPUTERS

Xavier Álvarez-Farré[1*], Andrey Gorobets[2], F. Xavier Trias[1] and Assensi Oliva[1]

[1] Technical University of Catalonia,
Heat and Mass Transfer Technological Center,
Carrer Colom 11, 08222 Terrassa (Barcelona), Spain
[2] Keldysh Institute of Applied Mathematics,
Miusskaya Sq. 4, 125047 Moscow, Russia

Key Words: High-Performance Computing, SpMV, Heterogeneous computing, NUMA-aware, Parallel CFD

Large sparse matrices often appear when numerically solving partial differential equations. Hence, the sparse matrix-vector product (SpMV) is a widely-used operation in the scientific computing community, and its implementation for hybrid supercomputers stirs a great deal of interest [1, 2]. The heterogeneous execution of the SpMV kernel leads to many different data-management and computing operations. The `halo` update, which is an expensive operation that can critically affect the performance and scalability, especially on hybrid supercomputers, must be concluded before computing the `interface` elements. Therefore, efficient execution strategies are required to minimise the overhead of communications. Roughly, heterogeneous execution strategies aim at handling computations and communications on multiple hardware devices through multithread parallel regions with separate kernel queues. The approach based on nested multithreading [3] hinders the efficient utilisation of multi-socket and non-uniform memory access (NUMA) architectures. In this work, we present a new approach based on flat multithreading which effectively increases the performance of the CPU side on such configurations, not prejudicing other devices.

REFERENCES

[1] W. Yang, K. Li, and K. Li, "A hybrid computing method of SpMV on CPU–GPU heterogeneous computing systems," *Journal of Parallel and Distributed Computing*, vol. 104, pp. 49–60, jun 2017.

[2] X. Álvarez, A. Gorobets, F. Trias, R. Borrell, and G. Oyarzun, "HPC^2–A fully-portable, algebra-based framework for heterogeneous computing. Application to CFD," *Computers & Fluids*, vol. 173, pp. 285–292, sep 2018.

[3] X. Álvarez, A. Gorobets, and F. X. Trias, "Strategies for the heterogeneous execution of large-scale simulations on hybrid supercomputers," in *7th European Conference on Computational Fluid Dynamics*, 2018.